COMPARISON OF EXON 5 SEQUENCES FROM 35 CLASS I GENES OF THE BALB/c MOUSE

By KURT A. BRORSON, STEPHEN W. HUNT III, TIM HUNKAPILLER, Y. HENRY SUN, HILDE CHEROUTRE, DEBORAH A. NICKERSON, AND LEROY HOOD

From the Division of Biology, California Institute of Technology, Pasadena, California 91125

The mouse class I MHC molecules are structurally related 45-kD cell surface glycoproteins that associate noncovalently with β_2-microglobulin, a 12-kD polypeptide (1). Class I molecules can be divided into two groups on the basis of their pattern of expression and their function. The transplantation antigens, H-2K, H-2D, and H-2L, are expressed on most somatic cells and present viral antigens to CTLs (2). The other group, the nonclassical class I molecules, exhibit a generally more restricted tissue distribution and are probably not involved in antigen presentation (3–7).

The BALB/c mouse has at least 35 class I genes that map to five genetic loci: *K*, *D*, *Qa*, *Tla*, and *Hmt* (8–11; Fig. 1). The classical transplantation antigen genes, K^d, D^d, and L^d map to the *K* and *D* loci, as do four other class I genes: $K2^d$, $D2^d$, $D3^d$, and $D4^d$ (12, 13). The *Qa* and *Tla* loci together contain 28 known class I genes, including some shown to encode nonclassical class I molecules (6, 14–16). In BALB/c mice, the eight *Qa* region genes are named $Q1^d$, $Q2^d$, $Q4^d$, $Q5^d$, $Q6^d$, $Q7^d$, $Q8/9^d$, and $Q10^d$, and the 19 *Tla* region genes are named $T1^c$ through $T18^c$ and 37^c. The newly described *Hmt* region contains at least three class I genes (10), including the *Thy-19.4* gene (11), which is included in this study.

Class I genes contain 6–8 exons (14, 17, 18). Exon 1 encodes a hydrophobic leader segment that is proposed to assist in the transport of the molecule to the cell surface and is cleaved post-translationally (19). Exons 2, 3, and 4 each encode the three 90-amino acid external domains: $\alpha 1$, $\alpha 2$, and $\alpha 3$. A short external connecting peptide, as well as the transmembrane domain and part of the cytoplasmic segment that includes charged anchoring residues, are encoded by exon 5 (Fig. 2). Exons 6, 7, and 8 encode the remainder of the cytoplasmic domain. Analysis of exon 5 sequences shows that they are generally not conserved for direct sequence similarity, but rather for maintaining hydrophobicity in the transmembrane stretch that they encode (20). Certain class I gene products, like those of the $Q4^d$ and $Q10^d$ genes,

This work was supported by National Institute of Health Grant AI-17565. S. W. Hunt is a Special Fellow of the Leukemia Society of America. D. Nickerson was a visiting associate from the University of South Florida, Tampa, FL. S. W. Hunt's present address is the Division of Rheumatology and Immunology, The University of North Carolina, Chapel Hill, NC 27599. Y. H. Sun's present address is the Institute of Molecular Biology, Academia Sinica, Nankang, Taipei 11529, Taiwan, Republic of China. H. Cheroutre's present address is the Department of Microbiology and Immunology, University of California, Los Angeles, CA 90024.

Address correspondence to K. A. Brorson, Division of Biology-147-75, California Institute of Technology, Pasadena, CA 91125.

are secreted and do not maintain a hydrophobic transmembrane domain. The $Q7^d$ gene product, the Qa-2 antigen, has a typical hydrophobic transmembrane domain and charged anchor residues, yet is linked to the cell surface via a phosphatidylinositol linkage (15). The transmembrane domain of the Qa-2 molecule is proposed to be cleaved before expression on the cell surface.

This report compares the exon 5 DNA sequences of the 35 known class I genes of the BALB/c mouse. Such a comparison can reveal which of these exons can encode a hydrophobic transmembrane, and whether the putative gene product could be membrane bound or secreted. Whereas the structure of the external α1, α2, and α3 domains has been resolved for at least one human class I antigen (21), no direct structural data exists for the transmembrane domains for the class I molecules. Therefore, an analysis of the predicted amino acid sequences of these genes could reveal what amino acid sequence and structural considerations are important for the function of the transmembrane domains. Analysis of the sequences reveals that, in spite of extensive nucleotide sequence variation, only four class I gene fifth exons, those from the $Q10^d$, $T5^c$, $T11^c$, and $T12^c$ genes, have frame shifts or stop codons that terminate their translation and prevent them from encoding a domain that is hydrophobic and long enough to span a lipid bilayer. Of the remaining fifth exons, 27 can encode membrane-spanning domains that resemble those of the classical transplantation antigens in that they can be divided into a proline-rich connecting peptide, a transmembrane segment, and a cytoplasmic segment with anchoring basic residues. In addition, hydrophobic moment analysis of the predicted transmembrane domains reveals that several, including those of the Qa-2 and TL antigens, are sufficiently amphipathic to promote intramembrane protein interactions. The conservation of the ability to encode a potentially functional transmembrane domain in the majority of the fifth exons suggests that selective pressure exists on them to remain functional, possibly because the majority of class I genes, including the divergent ones, are functionally important.

Materials and Methods

Sequencing of Transmembrane Exons. Individual class I genes or gene fragments were cloned from BALB/c MHC class I cosmids (8) into M13mp18- or pUC18-derived vectors. DNA sequencing was performed using the dideoxynucleotide chain termination method (22). Sequencing was primed with an oligonucleotide (5' ACCTTCCAGAAGTGGGCA 3') derived from a conserved area of the fourth exon of the L^d gene (23). This primer was chosen because the same sequence occurs in the fourth exon of several divergent class I genes, including the $H\text{-}2K^d$, D^d, L^d, $T13^c$ and $Q7^d$ (24) genes, and hence, is presumably highly conserved in most class I genes. Since exon 5 is generally ~120 nucleotides long and 210 nucleotides downstream of the primer, it was possible to determine the complete exon 5 sequence of all unpublished genes on one strand with one set of sequencing reactions. Exon 5 sequences that could not be directly aligned with previously reported sequences were also sequenced on the opposite strand using complementary oligonucleotide primers derived from intron 5 sequence. The fifth exons that can be aligned and were not sequenced on two strands were those from the $K2^d$, $Q1^d$, $Q4^d$, $Q5^d$, $Q6^d$, $Q10^d$, and $T3^c$ genes.

Sequence Alignments and Comparisons. Sequence alignments were performed using the method of Needleman and Wunsch (25), which inserts gaps into one or the other of the sequences in a pairwise comparison to maximize the similarity between the two sequences. In the percentage sequence similarity calculation, a gap of any size is counted as one mismatch, whereas unmatched sequences at either end are not counted.

After alignment, pairs of sequences were analyzed at each position for possible and observed silent and replacement substitutions (26). A single base change that does not change the predicted translation of coding region sequence is considered to be silent, while one that does change the predicted translation is considered to be a replacement. Each possible pairing of aligned sequences was analyzed, and substitutions were totaled for each category.

Analysis of Translated Exon 5 Sequences. The translated exon 5 sequences were analyzed by an algorithm that calculates the hydrophobicity of 21-amino acid stretches of the sequence (27). The hydrophobicity values of individual amino acids are taken from a consensus scale adapted from five separate hydrophobicity measurements (28). The method calculates which 21-amino acid stretch has the highest hydrophobicity value, thereby predicting which segment, if any, best defines the transmembrane domain.

Within the predicted 21-amino acid transmembrane segment, the hydrophobic moments, a measure of amphipathicity, of 11-amino acid stretches were calculated using the equation of Eisenberg et al. (27). The highest hydrophobic moment value for each transmembrane was plotted against the hydrophobicity value for the corresponding 11-amino acid stretch. The empirically defined area of the graph in which the point falls predicts where the predicted helix is likely to be found relative to the membrane, and whether it resides in the membrane alone or in association with another protein.

Results and Discussion

Exon 5 Sequences and Groups. The DNA sequences of the fifth exons of 35 BALB/c class I genes are shown in Fig. 3. In most cases, intron 4 and a portion of intron 5 are also included. The sequences were obtained from subclones of the BALB/c cosmids in this study, or from published sequencies. In most cases, exon 5 is identified by nucleotide similarity to known fifth exons, while in the cases of the *T7*^{c}, *T15*^{c}, and *Thy-19.4* genes, exon 5 is identified by the hydrophobicity of the translated amino acids, and relative position 3′ of exon 4. Exon boundaries are identified by comparison to class I genes for which spliced cDNA clones have been isolated (29, 30, Hunt, S., K. Brorson, H. Cheroutre, and L. Hood, manuscript in preparation), and by position of consensus splice sites. Donor splice sequences are not found in the *T4*^{c}, *T5*^{c}, *T7*^{c}, and *T15*^{c} fifth exons. The fifth exons of the *T7*^{c} and *T15*^{c} genes are interrupted by a B1 short interspersed repetitive element (31) after 143 bp, while those of the *T4*^{c} and *T5*^{c} genes are similar to other fifth exons for the first 46 and 58 bp, respectively, but contain nonhomologous sequences beyond what appears to be a

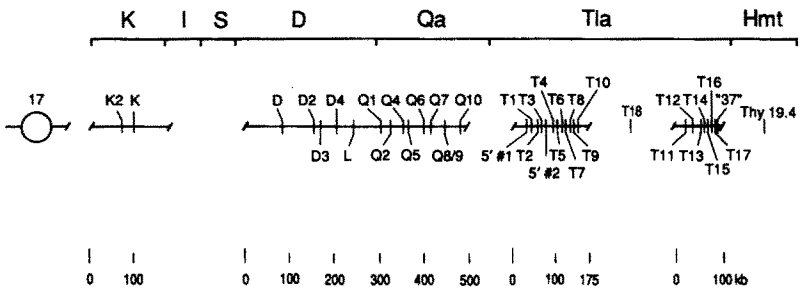


FIGURE 1. Map of class I genes in the BALB/c MHC. Class I genes map to the *K*, *D*, *Qa*, *Tla*, and *Hmt* regions. The *I* and *S* regions contain class II and complement genes, respectively. The order of the *Tla* region gene clusters is unknown, as is the distance between the *K*, *D*, *Tla*, and *Hmt* regions. The upper line represents the genetic map and the gene clusters are indicated below.

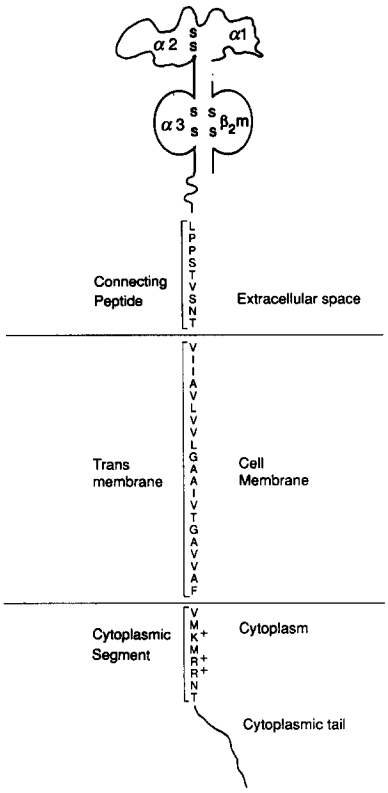


FIGURE 2. Model of typical class I transmembrane domain, H-2K^d. Connecting, transmembrane, and cytoplasmic segments are shown in which environment they are predicted to reside by the criteria used in this report (27).

recombinational or gene conversion boundary. The predicted reading frames are identified by the hydrophobicity of the translated amino acids, and by conformity to the reading frame established by the fourth exon.

The fifth exon sequences are assigned to the same group if they share at least 75% similarity with each other (32). The exon five sequences can be assigned to seven nonoverlapping groups. The largest group includes all of the *H-2* and *Qa* loci genes, and in addition, several even numbered *Tla* region genes: *T4*^{c}, *T6*^{c}, *T8*^{c}, *T10*^{c}, *T14*^{c}, *T16*^{c}, and *37*^{c}. A second group includes the *T1*^{c}, *T3*^{c}, *T11*^{c}, and *T13*^{c} genes, while a third includes the *T2*^{c}, *T5*^{c}, and *T12*^{c} genes. Finally, the *T9*^{c}/*T17*^{c} and *T7*^{c}/*T15*^{c} gene pairs form two additional groups, while the *T18*^{c} and *Thy-19.4* genes form two additional single gene groups. Consensus sequences are derived for each group based on the most frequent nucleotide used at each position.

Nucleotide sequence similarity among members of each group ranges from 73 to 99% (Table I). No two members of any group are exactly identical, making the exon 5 sequences diagnostic for the identification of BALB/c class I genes. Among members of each group, several types of mutational events have occurred subsequent to the duplications that created them, including nucleotide substitutions and short deletions. In addition, the exon 5 of the *Q1*^{d} gene has an extra 18 bp that matches 15 of the 18 nucleotides immediately following it, and thus, it is probably the product of an internal sequence duplication. Interestingly, an 18-bp insertion that matches

TABLE I

Percent Similarity of Fifth Exons within Groups

		Similarity of exons																			
Group	Exons	K^d	$K2^d$	D^d	$D2^d$	$D3^d$	$D4^d$	L^d	$Q1^d$	$Q2^d$	$Q4^d$	$Q5^d$	$Q6^d$	$Q7^d$	$Q8/9^d$	$Q10^d$	$T4^c$	$T6^c$	$T8^c$	$T10^c$	$T14^c$	$T16^c$
1	$K2^d$	86																			
	D^d	82	84																		
	$D2^d$	82	85	82																	
	$D3^d$	85	85	85	87																
	$D4^d$	87	90	85	85	85															
	L^d	82	85	85	85	88	85														
	$Q1^d$	85	85	85	85	91	87	85													
	$Q2^d$	89	89	86	88	91	90	90	91												
	$Q4^d$	84	86	78	85	87	86	88	85	88											
	$Q5^d$	83	84	98	83	86	85	85	85	87	79										
	$Q6^d$	83	84	78	84	86	86	87	83	87	97	78									
	$Q7^d$	83	86	82	88	90	86	87	88	90	87	83	86								
	$Q8/9^d$	83	85	81	87	89	86	86	87	84	86	82	87	99							
	$Q10^d$	73	78	76	82	79	79	84	78	8	80	77	79	81	80						
	$T4^c$	89	89	78	80	87	87	85	89	93	85	78	83	83	80	76					
	$T6^c$	81	84	81	81	83	86	83	88	88	81	82	80	81	80	75	93				
	$T8^c$	86	87	86	83	86	86	86	85	89	81	86	81	83	82	79	87	85			
	$T10^c$	83	85	82	85	85	83	83	87	89	80	83	79	84	83	76	85	81	82		
	$T14^c$	82	85	82	82	84	87	85	88	89	84	83	83	82	81	76	93	95	86	82	
	$T16^c$	87	88	87	84	87	87	87	86	90	82	87	81	84	83	80	89	86	99	83	87
	37^c	88	88	87	86	92	89	87	92	94	89	88	88	88	87	79	93	90	89	88	91
		$T1^c$	$T3^c$	$T11^c$																	
2	$T3^c$	90																			
	$T11^c$	94	93																		
	$T13^c$	91	98	95																	
		$T2^c$	$T5^c$																		
3	$T5^c$	95																			
	$T12^c$	98	91																		
4	$T7^c$ vs. $T15^c$	98																			
5	$T9^c$ vs. $T17^c$	99																			

In the comparisons, only the first 58 bp of the $T5^c$ fifth exon and 46 bp of the $T4^c$ fifth exon are used. The full lengths of the fifth exons are compared between all of the other genes.

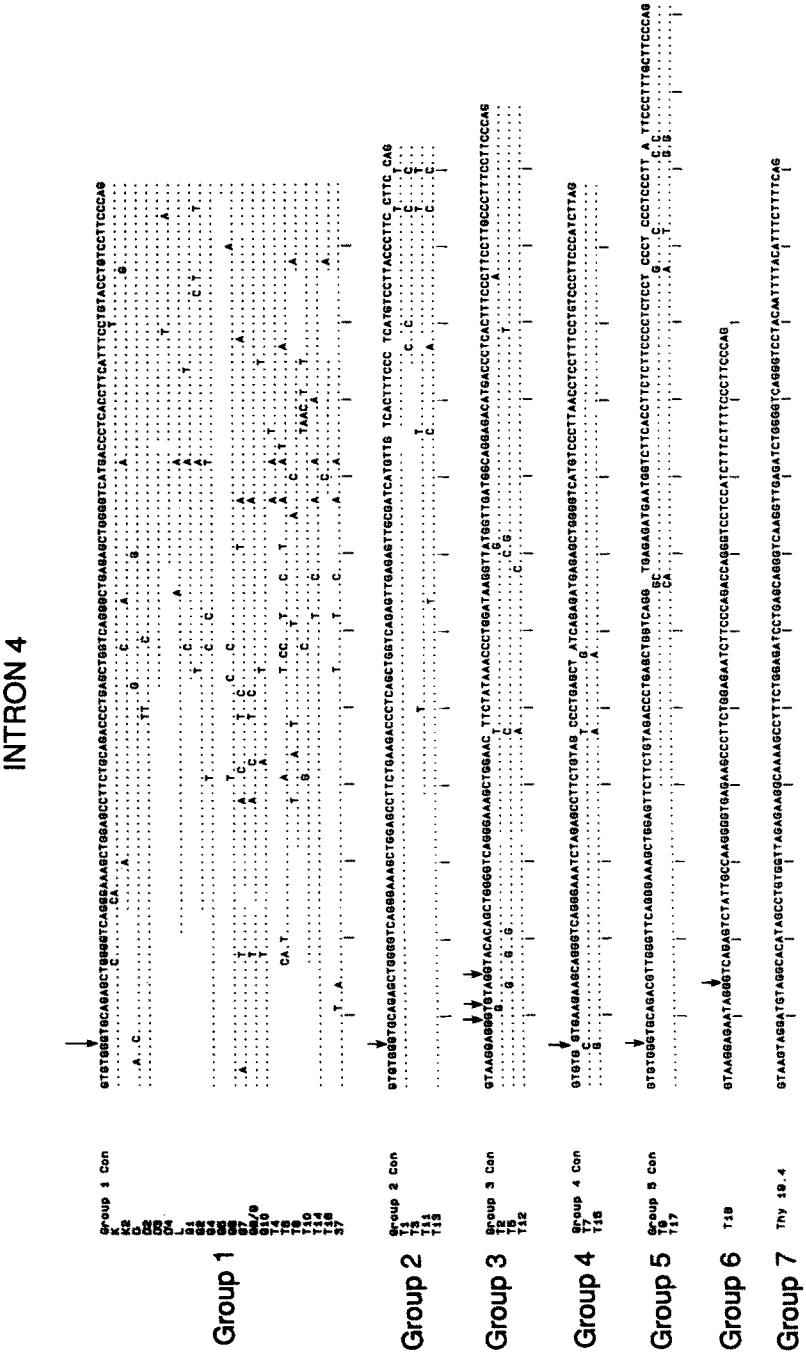
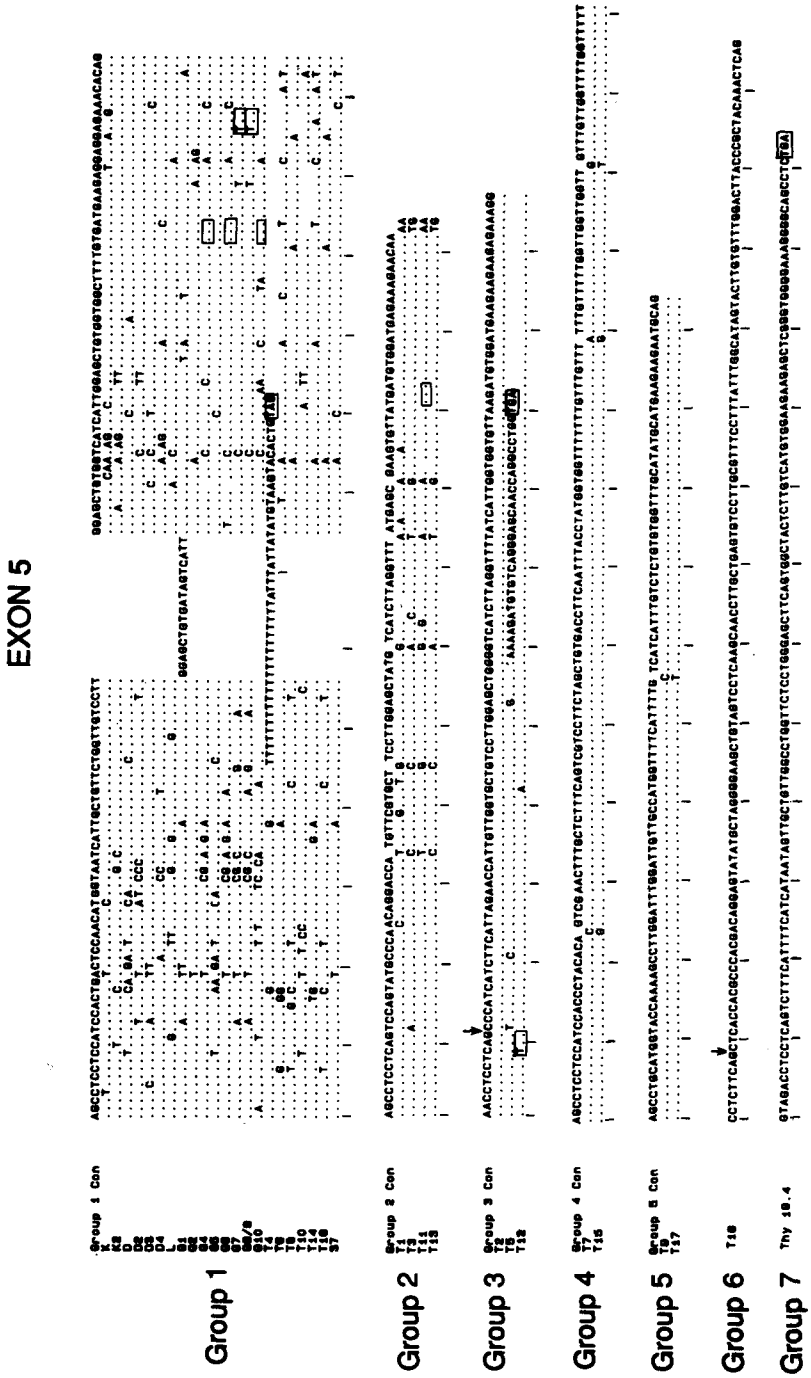


FIGURE 3. The DNA sequence of the fifth exons of the 35 BALB/c class I genes. Genes that were sequenced on one strand are the *K2*^{d}, *Q1*^{d}, *Q4*^{d}, *Q5*^{d}, *Q10*^{d}, and *T3*^{c} genes. All others were sequenced on both strands or have been previously published: *L*^{d} (22); *K*^{d} (17); *T13*^{c} (13); *D*^{d} (59); *37*^{c} (60); *Thy-19.4*^{c} (11); *T1*^{c} (61); *D2*^{d} (62); *Q7*^{d}, *T9*^{c}, *T17*^{c}, *T18*^{c}, *D3*^{d}, *D4*^{d} (Hunt et al., manuscript in preparation); *Q5*^{d}, *Q6*^{d}, and *Q8/9*^{d} (I. Stroynowski, personal communication); *Q10*^{d} (N. Ulker, personal communication). Intron 4 and 5 sequences are included in most cases. In-frame translation termination codons found in the fifth exons of the *Q4*^{d}, *Q6*^{d}, *Q7*^{d}, *Q8/9*^{d}, *Q10*^{d}, *T4*^{c},



T5ᶜ, *T7ᶜ*, *T11ᶜ*, *T12ᶜ*, *T15ᶜ*, and *Thy-19.4* genes are boxed. In the cases of the *T4ᶜ*, *T5ᶜ*, *T7ᶜ*, and *T15ᶜ* genes, the separation between exon 5 and intron 5 is arbitrary. In the *T4ᶜ* and *T5ᶜ* genes, the end of exon 5 is defined as the in-frame stop codon where translation terminates, while in the *T7ᶜ* and *T15ᶜ* genes, the B1 repeat element is arbitrarily included with the other intron 5 sequences. Possible alternative splice signals are indicated by arrows above the consensus sequences. These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession numbers X16197 through X16223. Figure continued on following page.

FIGURE 3. Continued. Legend on previous page.

the $Q1^d$ insertion in 13 of 18 nucleotides is also found in the same position in a rat class I gene (33; J. Howard, personal communication). Since the inserted sequence is about as similar to the rat sequence as it is to the 18-bp sequence following it, it is possible that this duplication event occurred before mouse/rat divergence. Alternatively, since this appears to be a single mutational event in both species, it is possible that the duplications occurred independently. Since the insertion in the $Q1^d$ exon 5 does not match precisely either the sequences immediately following it, or the homologous insertion in the rat fifth exon, it is unclear which of these two possibilities is the case.

Comparison of group consensus sequences reveals that some groups are related while others appear not to be (Table II). Groups 2 and 3 are ~73% similar to each other. Their members are different enough from each other to be classified as distinct groups based on the criteria of this report, but they are clearly evolutionarily related. The other groups are possibly related to each other since some pairings share as much as 54% similarity. Unlike the similarity between groups 2 and 3, it is unclear if 32–54% sequence similarity between these groups is the result of divergent evolution from an ancestral exon 5, or rather of convergent evolution of unrelated transmembrane exons. Codon usage that is restricted to maintain hydrophobicity of the translation could result in unrelated sequences attaining greater than random similarity. Thus, it is conceivable that the exon 5 sequences have multiple origins as the result of exon shuffling or de novo generation, and share similarity because of convergent evolution. This is most possible for the $T18^c$ fifth exon since it shares only 32–43% similarity to all of the other groups.

The existence of variation in the transmembrane exons in the class I gene family argues that their most important sequence consideration is the retention of a hydrophobic translation (20, 34). Extensive variation occurs in transmembrane domains when their only function is to anchor a protein to a membrane. To test whether selective pressure exists for the fifth exon sequences to retain their translation, synonymous and nonsynonymous mutation frequencies were determined for the group 1, 2, and 3 fifth exons (Table III). Since the members of groups 2 and 3 can be aligned,

TABLE II

Percent Similarity between Consensus Sequences of Fifth Exon Groups

	Similarity of exons					
Group	G1	G2	G3	G4	G5	G6
			%			
2	54					
3	53	73				
4	50	43	43			
5	41	38	48	46		
6	32	41	33	36	43	
7	48	38	49	40	42	37

Before the percent similarity calculation, the fifth exon sequences were aligned with gaps to maximize the result. Groups 1–7 consensus sequences are abbreviated as G1 through G7.

TABLE III

Replacement and Silent Site Mutation Frequencies in Exon 5 and Intron 4

	Replacement	Silent	Possible	Observed
Exon 5 Group 1	2,766/17,249 (16)	765/6,133 (12)	2.81	3.62
Exon 5 Groups 2 and 3 (aligned)	202/1,259 (16)	40/361 (11)	3.49	5.05
Intron 4 (Exon 4 read-through) Group 1	1,264/15,319 (8)	594/5,951 (10)	2.57	2.13
Intron 4 Groups 2 and 3 (aligned)	217/1,568 (14)	68/496 (14)	3.16	3.19

Frequencies are expressed both as a fraction of observed changes over possible changes, and as a percentage (in parentheses). Possible and observed replacement/silent ratios are also shown. The *T4*^{c} and *T5*^{c} fifth exons were omitted from this analysis since their 3′ portions were created by recombination events, not duplication and point mutation.

they were pooled to maximize the number of sites tested. As a contrast, synonymous and nonsynonymous mutation frequencies were also determined in the exon 4 read-through frame of intron 4. If selective pressure is exerted on a coding region sequence, the frequency of silent mutations is predicted to be higher than that of replacement mutations. As expected, silent and replacement mutations are approximately equivalent in the intron 4 sequences. However, in the fifth exons of groups 1, 2, and 3, replacement mutations have a higher frequency than silent mutations. This suggests that there is little selective pressure to maintain their protein encoding sequences other than for hydrophobicity, although the variation in the putative *Tla* region gene-encoded transmembranes may have evolved because they perform specialized functions that are different than those of the transplantation antigens.

In contrast to the fifth exon sequences, almost all exon 4 sequences are at least 80% similar to each other (Hunt, S., K. Brorson, H. Cheroutre, and L. Hood, manuscript in preparation, K. Brorson unpublished observations), supporting the concept that all of the class I genes evolved from a common ancester (35). It is interesting that the highly conserved fourth exons and the highly divergent fifth exons are separated only by a 120-nucleotide intron. Dot matrix identity plots between group consensus sequences (Fig. 4) reveal that, between groups, intron 4 is more conserved than exon 5, and that there are two general areas of conservation. One area is the splice acceptor site and the first ~10 bp of exon 5. The other area is the 5′ portion of intron 4, adjacent to the conserved exon 4. It is conserved among all of the genes except the *T18*^{c} gene, and the *Thy-19.4* gene, where only the middle of the intron is conserved. Intron 4 is also more conserved than exon 5 when compared among groups (Table III). However, it is unlikely that this reflects selective pressure for their conservation, as would occur if read-through translation from exon 4 is important since the silent mutation frequency of the read-through frame is approximately equal to the replacement frequency in group 1, as well as in groups 2 and 3. Instead, the

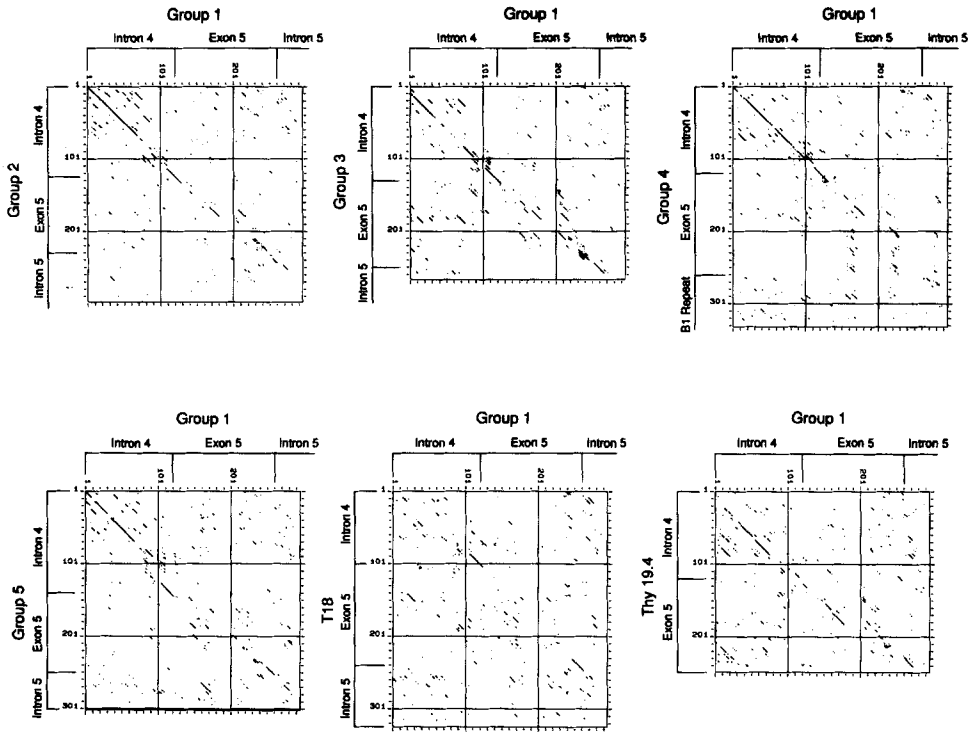


FIGURE 4. Dot matrix identity comparison of group 1 consensus sequences with groups 2–7 consensus sequences. Exon 5 boundaries are shown on the top and on the side. Each dot represents a six of eight nucleotide match between the sequences.

distinct breaks in similarity evident in the dot matrix identity plots suggest that the conservation in the fourth intron is a result of recombinational events that were involved in the evolution of the class I gene family. These recombination events could have included transmembrane exon shuffling or de novo generation events that created hybrid genes with similarity to classical class I genes in exon 4 and the 5′ portion of intron 4, but little or no similarity in exon 5 and the rest of intron 4. Alternatively, it is suggested that short introns in class I genes are generally more conserved than the interior portions of long introns because proposed recombination events that transfer exons between class I genes could often extend beyond the end of the exons into a portion of the surrounding introns (36). It is conceivable that the 5′ portions of the fourth introns are generally more conserved than the 3′ portions because such DNA segment exchange events could occur more often between fourth exons than fifth exons (14).

The exon 5 sequence data can be used to support models for the evolution of specific groups of class I genes. It is proposed that the $Q4^b$-$Q10^b$ genes in the C57BL/10 mouse resulted from duplications of a primordial *Qa* gene pair, with the even- and odd-numbered genes derived from one or the other of the primordial genes (37, 38). The *Q8* and *Q9* genes were subsequently fused in an unequal crossover event to form

the hybrid gene *Q8/9*^{d} of BALB/c mice (37). The exon 5 sequence data supports this model since exon 5 in the *Q4*^{d} and *Q6*^{d} genes share 97% similarity and a single nucleotide deletion causing a frame shift at nucleotide 62. In addition, the *Q7*^{d} and *Q8/9*^{d} genes are 99% similar to each other in exon 5. However, the *Q5*^{d} gene is 98% similar to the *D*^{d} gene in exon 5, suggesting that it had undergone a DNA segment exchange event from that gene.

The two gene clusters, *T1*^{c}*-T10*^{c} and *T11*^{c}*-37*^{c} (Fig. 1), of the *Tla* region of the BALB/c mouse, are proposed to have resulted from a duplication of an entire block of genes (14). The exon 5 sequence groups define gene pairs with representatives in the same order on both clusters. These gene pairs are *T1*^{c}*/T11*^{c}, *T2*^{c}*/T12*^{c}, *T3*^{c}*/T13*^{c}, *T6*^{c}*/T14*^{c}, *T7*^{c}*/T15*^{c}, *T8*^{c}*/T16*^{c}, *T9*^{c}*/T17*^{c}, and *T10*^{c}*/37*^{c}. The placement of these pairs in a specific order in both clusters supports the cluster duplication model. Because of similarities in restriction enzyme site patterns, the *T4*^{c} and *T5*^{c} genes are proposed to have been created in a duplication of a pair of genes that also produced the *T6*^{c} and *T7*^{c} genes (14). However, the exon 5 sequence data does not support this contention since the *T5*^{c} exon 5 does not resemble the *T7*^{c} exon 5, but instead is 95% similar to the *T2*^{c} exon 5 in the first 58 nucleotides. Beyond that point, it does not resemble any other exon 5 sequence. In addition, exon 5 in the *T4*^{c} gene is 93% similar to that of the *T6*^{c} gene in the first 46 nucleotides, after which is a 21-bp polythymidine tract followed by a nonhomologous sequence. Since both the *T4*^{c} and *T5*^{c} exon 5 sequences are interrupted by nonhomologous sequences, it is likely that instead of being created as a block duplication of the *T6*^{c} and *T7*^{c} genes, the *T4*^{c} and *T5*^{c} genes are partial class I genes that were duplicated separately from distinct sources. The *T5*^{c} gene is probably a partially duplicated *T2*^{c} or *T12*^{c} gene, while the *T4*^{c} gene is probably a partially duplicated group 1 class I gene.

Splice Junction Sequences. The putative acceptor splice junction sequences of 35 and donor sequences of 30 of the class I fifth exons are shown in Fig. 5. The fifth exons of the *T4*^{c}, *T5*^{c}, *T7*^{c}, and *T15*^{c} genes do not have donor splice sequences; probably because they were eliminated by recombination or repetitive element integration events during their evolution. In addition, since *Thy-19.4* transcripts do not splice exon 5 to any 3′ exons (11), it is also excluded from the donor sequence figure. All 35 acceptor sequences have polypyrimidine tracts of 16–58 bp in length, followed by an AG dinucleotide. Since splicing invariably occurs after an AG dinucleotide in eukaryotic genes (39), and polypyrimidine tracts in acceptor splice signals are generally >11 bp in length, all 35 acceptor splice sequences appear functional. Similarly, all of the donor splice sequences match the consensus sequence ($^{C}_{A}$AG/GT$^{A}_{G}$AGT) with at least six of nine nucleotides and have the invariant GT dinucleotide at the immediate splice junction. Since donor splice sequences only need to match the consensus sequence in as little as five of the nine nucleotides to be functional (40), and since all of the class I donor sequences have the invariant GT dinucleotide found in all eukaryotic donor sequences (39), all 30 of the class I donor sequences also appear to be functional. Since none of the splice sequences in the 35 class I fifth exons appear abnormal, it is unlikely that any of the fifth exons will be nonfunctional because of splicing abnormalities.

In addition to donor and acceptor splice sequences homologous to those in previously characterized genes, several possible alternative splice sites can be identified (Fig. 3). These sites include in-frame donor sequences in members of groups 1–5

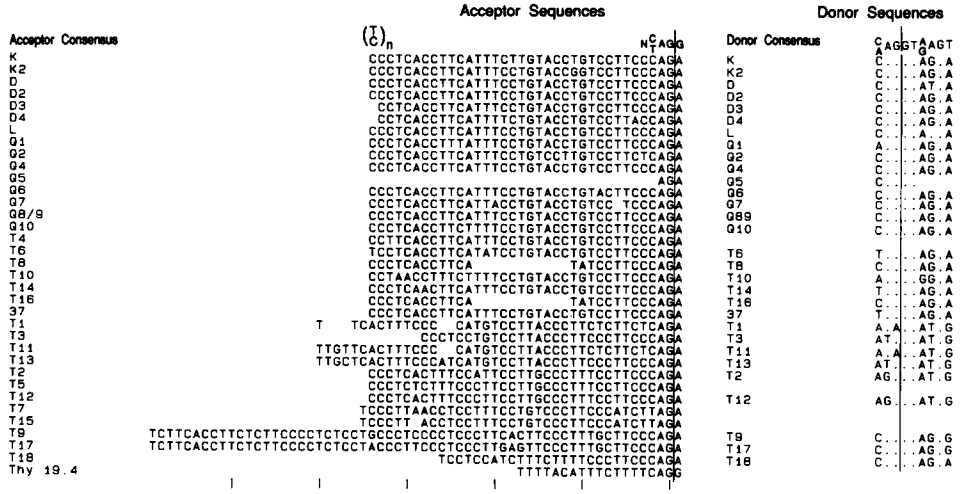


FIGURE 5. The acceptor and donor splice sequences of BALB/c class I genes. The consensus sequence (Con) appears above the compiled sequences, and vertical lines indicate where splicing is predicted to occur.

in intron 4 near the junction with exon 4. If a class I transcript splices at these sites, between two and five amino acids would be added to the α3 domain of its translation product. In the *T2*^{c}, *T5*^{c}, *T12*^{c}, and *T18*^{c} genes, there are additional possible alternative donor sites in intron 4 that generate a different frame than that identified in cDNA transcripts. However, transcripts that splice these possible alternative donor sites to possible alternative acceptor sites in the 5′ portion of exon 5 would place the hydrophobic translation of exon 5 in the same reading frame as exon 4. The translation would be slightly longer in exon 4 and slightly shorter in exon 5. The translation of the *T12*^{c} fifth exon terminates two amino acids after the acceptor splice signal homologous to those characterized in other class I genes. However, a *T12*^{c} transcript could encode a hydrophobic transmembrane if spliced at the possible alternative splice sites. Finally, there are donor signals in the fifth exon of the *T4*^{c} gene. However, since these sites were probably introduced to this gene by a recombination event, it is unclear if they actually evolved to splice the *T4*^{c} fifth exon to a 3′ exon.

Analysis of Predicted Sequence. Translation of the DNA sequences reveals that, with the exception of the *Q10*^{d}, *T5*^{c}, *T11*^{c}, and *T12*^{c} genes, each class I gene has an open reading frame in exon 5, whose translation is potentially hydrophobic and long enough to span a lipid bilayer (27; Fig. 6). Thus, each of these 31 fifth exon–encoded amino acid sequences can be divided into connecting, transmembrane and cytoplasmic segments. In this study the transmembrane segment is arbitrarily defined as the most hydrophobic 21 amino acids of the fifth exon translation, since that is the chain length required to form an α helix that can completely span a lipid bilayer (41). Analysis of previously characterized membrane-bound proteins reveals that transmembrane domains range in length from 20 to 28 amino acids (42). Since the choice of 21 amino acids is arbitrary, it is important to note that it is possible that in the actual gene

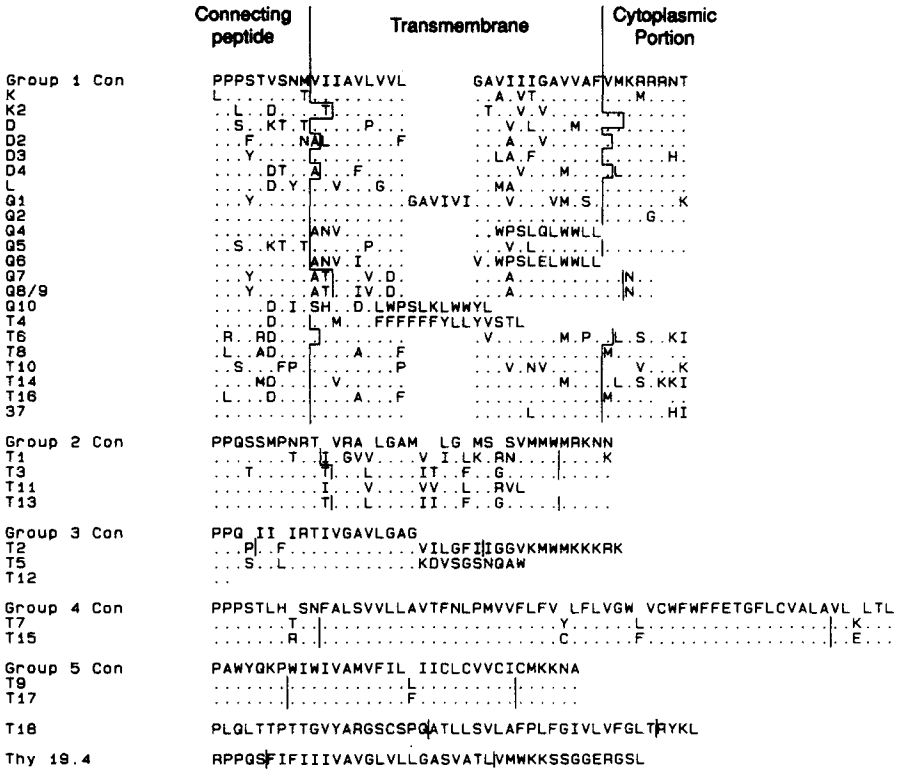


FIGURE 6. The translations of the exon 5 sequences. Where they exist, borders between predicted connecting, transmembrane, and cytoplasmic segments are indicated by vertical lines between amino acids. Frameshifts present in the fifth exons of the *Q4*^{d}, *Q6*^{d}, and *Q10*^{d} genes cause a portion of their exon 5 translation to be nonhomologous to those of other class I genes in their group, while recombination events that occurred in the *T4*^{c} and *T5*^{c} fifth exons produced a similar result for their translations.

products some of the residues near the calculated borders may not reside in their predicted environment. The exceptions to the arbitrary assignment of 21 amino acids are the predicted Q1^d, T7^c, and T15^c transmembrane domains, which clearly have hydrophobic segments in excess of 27 amino acids. In the translated sequences, the putative connecting peptides vary from 4 to 20 amino acids in length. The putative cytoplasmic portions are between 4 and 13 amino acids in length, except in the Q4^d, Q6^d, T4^c, and T15^c transmembrane domains, where none can be identified.

Connecting Peptides. Analysis of the putative connecting peptides reveals that the amino acid usage is similar to that in the hinge regions of Igs (24; Table IV). Proline (28%) is the most commonly used amino acid. In addition, asparagine (8%) is also present in these segments. These amino acids tend to disrupt any helical structure that may form in the junction between the transmembrane and outer domains (43). In addition, serine and threonine predominate in the connecting peptide at 16 and 14%, respectively. These two amino acids with small polar hydroxyl side chains are common in exposed areas, and their presence is not predicted to contribute to or disrupt the formation of α helices (43). However, comparison of 31 proteins exhibiting

TABLE IV
Amino Acid Compositions of Predicted Connecting, Transmembrane, and Cytoplasmic Segments

Amino acid	Connecting segment	Transmembrane	Cytoplasmic portion
		%	
Acidic			
Aspartic acid (D)	2.5	0.3	0
Glutamic acid (E)	0	0.4	1.0
Basic			
Lysine (K)	1.4	0.1	19.0
Arginine (R)	2.8	0.6	25.1
Histidine (H)	0.7	0	1.0
Polar			
Glycine (G)	0.7	8.3	3.1
Asparagine (N)	8.2	0.9	10.8
Glutamine (Q)	3.2	0.1	0
Cysteine (C)	0.4	1.6	1.0
Serine (S)	16.0	1.7	3.1
Threonine (T)	14.2	1.6	6.7
Tyrosine (Y)	2.8	0.1	0.5
Nonpolar			
Alanine (A)	2.8	15.1	1.0
Valine (V)	4.6	26.5	7.7
Leucine (L)	2.8	12.9	5.6
Isoleucine (I)	0	14.2	2.1
Proline (P)	28.4	1.3	0
Phenylalanine (F)	0.7	7.8	0
Methionine (M)	7.1	3.5	11.3
Tryptophan (W)	0.7	2.8	1.0

The calculation reflects percent representation in a total of 31 connecting and transmembrane segments, and 27 cytoplasmic segments. The individual amino acids have been previously assigned to acidic, basic, polar, or nonpolar categories (58).

segment flexibility demonstrates that both serines and threonines tend to be concentrated in flexible segments (44). The serines and threonines in the connecting peptides could confer more flexibility to this segment.

The imposition of a flexible β-turn structure in the connecting peptide could facilitate stretching and pivoting at this segment in a manner similar to that in the hinge region of Igs. It is suggested that the freedom of movement of the two Ig Fab arms relative to the Fc stem results from proline-rich amino acid sequences within the hinge segment that favor flexibility (24, 45, 46). This freedom of movement of the Fab arms is believed to be important for the function of Igs (47). Similarly, in the case of transplantation antigens, freedom of movement in the connecting peptide could be important to facilitate interaction with the TCR. It is interesting that the $T18^c$ molecule is predicted to have a connecting peptide 20 amino acids in length. It would be twice as long as those in the transplantation antigens, but it is unclear if there is any significance to this difference.

Transmembrane Segments. In the predicted transmembrane segments, four hydrophobic amino acids dominate: valine (26%), alanine (15%), isoleucine (14%), and leucine (13%) (Table IV). Phenylalanine is present at an intermediate level of 8%, but the other three hydrophobic amino acids, tryptophan, methionine, and proline, each constitute 4% or less of the transmembrane amino acids. Proline (1%) may be absent from the transmembrane segments because it may tend to disrupt the α-helical structure assumed by the hydrophobic amino acids in their aliphatic environment (48, 49), although it has been suggested, on the basis of work with bacterial transmembrane domain deletion mutants, that transmembrane domains are not always completely α helical (50). Tryptophan (3%) and methionine (4%) are used less often in proteins in general, and their lower usage in the transmembrane domains could reflect this (51). In addition, tryptophan is suggested to be more hydrophilic than previously believed, and is represented infrequently in other transmembrane segments (52). Glycine (8%) is the only nonhydrophobic amino acid found in abundance in the transmembrane segments. Glycine has a very small slightly polar side chain that would probably not significantly decrease the hydrophobicity of the transmembrane segments. Interestingly, the putative $Q1^d$, $T7^c$, and $T15^c$ molecules are predicted to have hydrophobic transmembrane segments of between 27 and 49 amino acids. Although these genes have not been shown to encode proteins, if they did, it would be interesting to see how such long hydrophobic segments are accommodated in the membrane.

Proteins with transmembrane domains that interact with other proteins within the lipid bilayer, as class II MHC molecules are proposed to (53), are suggested to do so because they contain short stretches within their membrane spanning segment that are sufficiently amphipathic to promote such interactions (27). To test whether any of the class I transmembrane segments, as well as four BALB/c class II transmembrane segments, A^d_α, A^d_β, E^d_α, and E^d_β, could interact within the membrane with other proteins, the hydrophobic moment, a measure of amphipathicity, was calculated for 11-amino acid stretches within them. The length of 11 amino acids corresponds to approximately three turns of an α helix, which is believed to be the typical amphipathic segment size that interacts noncovalently with other proteins. For each transmembrane segment, the 11-amino acid stretch with the highest hydrophobic moment was determined, and that hydrophobic moment value was plotted against the hydrophobicity value for the 11-amino acid stretch (Fig. 7). Whether an 11-amino acid stretch is predicted to be sufficiently amphipathic to promote interactions within the membrane depends on which empirically defined area within the graph its plotted point falls (27).

This analysis reveals that the plotted points of all four class II transmembranes fall within or near the area defined as multimeric transmembrane. Since class II molecules are dimeric on the cell surface, it is proposed that the transmembrane's amphipathicity and amino acid sequence conservation are consistent with the hypothesis that they are dimeric within the lipid bilayer as well (53). On the other hand, the algorithm predicts that several class I transmembrane domains are not sufficiently amphipathic to be predicted to interact with other proteins within the membrane. The transplantation antigens, K^d, D^d, and L^d, are heterodimers with β_2-microglobulin, a small polypeptide with no membrane-spanning segment. Therefore, the prediction that they do not have amphipathic transmembrane segments is consis-

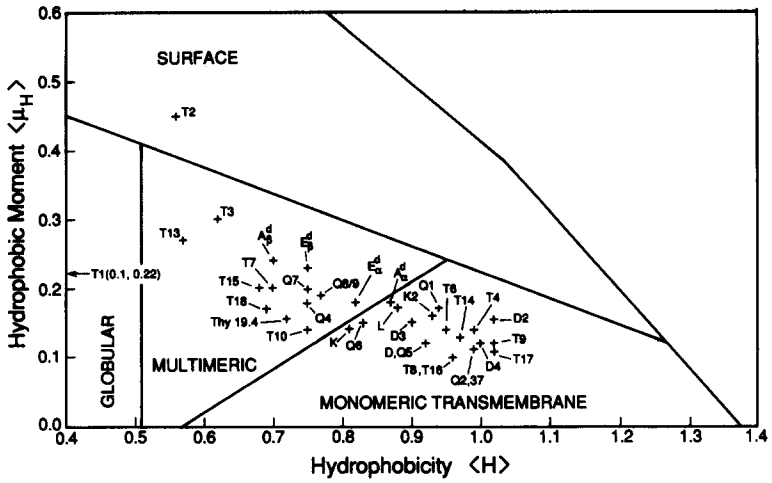


FIGURE 7. Hydrophobic moment plot. Hydrophobic moment (μ_H) is plotted against hydrophobicity (H) for 11-amino acid segments within each of 31 class I fifth exon translations, as well as the transmembrane domains of class II molecules A^d_α, A^d_β, E^d_α, and E^d_β. Each point plots within arbitrary areas labeled surface, globular, multimeric, or monomeric transmembrane. Although arbitrarily defined by Eisenberg et al. (27), 36 of 49 transmembrane segments originally used to define these regions were correctly plotted within the region of the graph that corresponded to their type.

tent with their probable monomericity within the membrane. In addition to the transplantation antigens, the putative molecules encoded by the majority of group 1 genes and the *T9*^{c} and *T17*^{c} genes are also predicted to be monomeric within the membrane. However, other putative class I transmembrane segments are predicted to be sufficiently amphipathic to associate within the membrane with other proteins. These include those predicted to be encoded by the *Q4*^{d}, *Q7*^{d}, *Q8/9*^{d}, *T3*^{c}, *T7*^{c}, *T10*^{c}, *T13*^{c}, *T15*^{c}, *T18*^{c}, and *Thy-19.4* genes. The *Q4*^{d} gene encodes a secreted class I product, and the *Q7*^{d} gene encodes the Qa-2 antigen, which is linked to the cell surface by a phosphatidylinositol linkage. The amphipathicity data is consistent with the hypothesis that during the processing or transport of these two molecules, intramembrane interactions occur with yet uncharacterized proteins. In addition, the putative products of several *Tla* region class I genes, including that of the *T13*^{c} gene, which encodes the TL^c antigen, are also predicted to interact with other proteins within the membrane. These *Tla* region genes, *T3*^{c}, *T7*^{c}, *T13*^{c}, *T15*^{c}, *T18*^{c}, and *Thy-19.4*, also have highly divergent fifth exons, suggesting that the molecules that these genes encode perform functions different than those of the classical class I molecules. The hydrophobic moment data is consistent with the hypothesis that their putative function may require intramembrane interactions with other molecules, possibly for the initiation of signaling cascades.

Cytoplasmic Segment. In the cytoplasmic portion, two basic amino acids predominate: arginine (25%) and lysine (19%). Basic amino acids are commonly found on the cytoplasmic side of transmembrane domains and are proposed to prevent the short cytoplasmic domain from being pulled through the hydrophobic lipid bilayer (42, 54). Histidine, a slightly basic amino acid, is not present in the class I anchor se-

quences. It is probably too weakly basic to serve in an anchor sequence. Also present in the cytoplasmic segments are methionine (11%), valine (8%), and asparagine (11%). It is unclear if there is any significance to the presence of these amino acids, although it is interesting that methionines and asparagines are clustered at both ends, but not at the center, of the transmembrane domains. Some of these residues may be spacers at the end of the domain and between the highly charged basic residues. Others are transmembrane segment amino acids included in cytoplasmic portion because of the arbitrary decision to limit the transmembrane segment to the 21 most hydrophobic amino acids.

Four of the 31 hydrophobic transmembrane domains do not have basic anchoring residues at the COOH-terminal end: $Q4^d$, $Q6^d$, $T4^c$, and $T15^c$. It is known that the *Q4*^{d} gene, like the *Q10*^{d} gene, encodes a secreted class I molecule (7, 55). The lack of an anchor sequence probably contributes to the fact that it is a secreted class I, in spite of its hydrophobic transmembrane segment. On the other hand, studies with *H-2L*^{d} gene mutants demonstrate that anchoring residues are not absolutely necessary for cell surface expression of class I glycoproteins (56). In addition, it is suggested that the $Q4^P$ molecule can exist on the cell surface of transfected cells (57). Clearly, the absence of anchoring residues can not universally be used as criteria for whether a class I molecule is secreted or membrane expressed. If there are products of the *Q6*^{d}, *T4*^{c}, and *T15*^{c} genes, it will be interesting to see whether they are secreted, membrane bound, or both. Interestingly, the putative transmembrane domains encoded by the *T7*^{c}, *T9*^{c}, and *T17*^{c} genes are predicted to end with only one or two basic anchoring residues, whereas the classical transplantation antigens are anchored by three to four basic residues (Fig. 6). If these genes can encode class I molecules, it will also be interesting to see if these molecules are anchored to the cell membrane as efficiently as the transplantation antigens. The $Q10^d$ molecule has neither a hydrophobic transmembrane nor anchoring residues (55), and it is known to be secreted (6).

Implication of Predicted Protein Sequences. This analysis reveals that almost all of the 35 class I genes have fifth exons that have open reading frames that could potentially encode a domain that is sufficiently hydrophobic and long to span a lipid bilayer, and hence by this criterion, appear to be functional. Only four of the 35 BALB/c class I gene fifth exons, those of the *Q10*^{d}, *T5*^{c}, *T11*^{c}, and *T12*^{c} genes, appear to be exceptions and have stop codons or frame shifts that prevent them from encoding a hydrophobic transmembrane domain. However, this does not necessarily imply that these four genes are pseudogenes, since at least the *Q10*^{d} gene encodes a presumably functional soluble class I molecule. Thus, based on the analysis of these sequences, it is not evident that any of these genes are pseudogenes. Of the remaining 31 class I genes, 27 have a fifth exon that could encode a domain similar to those of transplantation antigens in that it has both hinge-like connecting peptides and basic anchor amino acids at the appropriate ends of the hydrophobic stretch. Only the $Q4^d$, $Q6^d$, $T4^c$, and $T15^c$ transmembrane domains are exceptions by lacking basic anchor amino acids. Overall, the amino acid usage of these segments is appropriate for their predicted function. The hinge-like segments use amino acids expected to introduce β turns and segmental flexibility. The transmembrane segments consist of hydrophobic amino acids, while there are anchoring basic residues in the cytoplasmic segments. The maintenance of this motif is particularly striking because

of the extensive sequence divergence of the fifth exon groups. Since the majority of the fifth exons appear to be able to encode a functional transmembrane domain, it is unlikely that their divergence is merely a result of genetic drift in the absence of selective pressure. It is more likely that selective pressure exists to maintain them, suggesting that the majority of the class I genes, including the divergent ones, are functionally important. It could be speculated that the fifth exons have diverged from each other because the molecules that they encode have specialized functions other than antigen presentation to T cells. If the molecules that the divergent groups encode are involved in other functions, the fifth exons would still be selected for the ability to encode a transmembrane domain, but not one similar to those in restriction elements. Clearly, analysis of exon 5 sequences alone can only suggest what a particular class I gene can encode; further sequence and expression studies will be required to determine the extent of expression of class I genes.

Summary

DNA sequences of the fifth exon, which encodes the transmembrane domain, were determined for the BALB/c mouse class I MHC genes and used to study the relationships between them. Based on nucleotide sequence similarity, the exon 5 sequences can be divided into seven groups. Although most members within each group are at least 80% similar to each other, comparison between groups reveals that the groups share little similarity. However, in spite of the extensive variation of the fifth exon sequences, analysis of their predicted amino acid translations reveals that only four class I gene fifth exons have frameshifts or stop codons that terminate their translation and prevent them from encoding a domain that is both hydrophobic and long enough to span a lipid bilayer. Exactly 27 of the remaining fifth exons could encode a domain that is similar to those of the transplantation antigens in that it consists of a proline-rich connecting peptide, a transmembrane segment, and a cytoplasmic portion with membrane-anchoring basic residues. The conservation of this motif in the majority of the fifth exon translations in spite of extensive variation suggests that selective pressure exists for these exons to maintain their ability to encode a functional transmembrane domain, raising the possibility that many of the nonclassical class I genes encode functionally important products.

We thank Drs. I. Stroynowski, M. Zuniga, K. Fischer Lindahl, and J. Kobori for critically reviewing this manuscript; Dr. J. Howard for critical insights on exon 5 evolution and sharing unpublished rat exon 5 sequence data; and Mrs. C. Blagg for expert secretarial assistance.

Received for publication 19 June 1989 and in revised form 22 August 1989.

References

1. Silver, J., and L. Hood. 1974. Detergent solubilized H-2 alloantigen is associated with a small molecular weight polypeptide. *Nature (Lond.).* 249:764.
2. Zinkernagel, R. M., and P. C. Doherty. 1980. MHC-restricted cytotoxic T cells: studies on the role of polymorphic major transplantation antigens determining T-cell restriction specificity, function, and responsiveness. *Adv. Immunol.* 27:51.
3. Old, L. J., E. A. Boyse, and E. Stockert, E. 1963. Antigenic properties of experimental leukemias. I. Serological studies *in vitro* with spontaneous and radiation-induced leukemias. *J. Natl. Cancer Inst.* 31:977.

4. Flaherty, L. 1976. The *Tla* region of the mouse: identification of a new serologically defined locus, *Qa-2*. *Immunogenetics*. 3:533.
5. Stanton, T. H., and E. A. Boyse. 1976. A new serologically defined locus, *Qa-1*, in the *Tla* region of the mouse. *Immunogenetics*. 3:525.
6. Maloy, W., J. Coligan, Y. Barra, and G. Jay. 1984. Detection of a secreted form of the murine H-2 class I antigen with an antibody against its predicted carboxyl terminus. *Proc. Natl. Acad. Sci. USA*. 81:1216.
7. Robinson, P. J. 1985. Qb-1, a new class I polypeptide encoded by the *Qa* region of the mouse H-2 complex. *Immunogenetics*. 22:285.
8. Steinmetz, M., A. Winoto, K. Minard, and L. Hood. 1982. Clusters of genes encoding mouse transplantation antigens. *Cell*. 28:489.
9. Winoto, A., M. Steinmetz, and L. Hood. 1983. Genetic mapping in the major histocompatibility complex by restriction enzyme polymorphism: most mouse class I genes map to the *Tla* complex. *Proc. Natl. Acad. Sci. USA*. 80:3425.
10. Richards, C. S., M. Bucan, K. Brorson, M. Kiefer, S. Hunt, H. Lehrach, and K. Fischer Lindahl. Genetic and molecular mapping of the *Hmt* region of the mouse. *EMBO (Eur. Mol. Biol. Organ.) J.* In press.
11. Brorson, K., S. Richards, S. Hunt, H. Cheroutre, K. Fischer Lindahl, and L. Hood. 1989. Analysis of a new class I gene mapping to the *Hmt* region of the mouse. *Immunogenetics*. 30:273.
12. Goodenow, R., M. McMillan, M. Nicolson, B. Sher, K. Eakle, N. Davidson, and L. Hood. 1982. Identification of the class I genes of the mouse major histocompatibility complex by DNA-mediated gene transfer. *Nature (Lond.)*. 300:231.
13. Stephan, D., H. Sun, K. Fischer Lindahl, E. Meyer, G. Hämmerling, L. Hood, and M. Steinmetz. 1986. Organization and evolution of *D* region class I genes in the mouse major histocompatibility complex. *J. Exp. Med.* 163:1227.
14. Fisher, D. A., S. W. Hunt, and L. Hood. 1985. Structure of a gene encoding a murine thymus leukemia antigen, and organization of *Tla* genes in the BALB/c mouse. *J. Exp. Med.* 162:528.
15. Stroynowski, I., M. Soloski, M. Low, and L. Hood. 1987. A single gene encodes soluble and membrane-bound forms of the major histocompatibility Qa-2 antigen: anchoring of the product by a phospholipid tail. *Cell* 50:759.
16. Robinson, P., D. Bever, A. Mellor, and E. Weiss. 1988. Sequence of the mouse *Q4* class I gene and characterization of the gene product. *Immunogenetics*. 27:79.
17. Steinmetz, M., K. W. Moore, J. Frelinger, B. Sher, F. Shen, E. A. Boyse, and L. Hood. 1981. A pseudogene homologous to mouse transplantation antigens: transplantation antigens are encoded by eight exons that correlate with protein domains. *Cell*. 25:683.
18. Kvist, S., L. Roberts, and B. Dobberstein. 1983. Mouse histocompatibility genes: structure and organization of a K^d gene. *EMBO (Eur. Mol. Biol. Organ.) J.* 2:245.
19. Blobel, G., and B. Dobberstein. 1975. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane bound ribosomes of murine myeloma. *J. Cell Biol.* 67:835.
20. Uehara, H., J. Coligan, and S. Nathenson. 1981. Isolation and sequence analysis of the intramembranous hydrophobic segment of the H-$2K^b$ murine histocompatibility antigen. *Biochemistry*. 20:5936.
21. Bjorkman, P., M. Saper, B. Samraoui, W. Bennett, J. Strominger, and D. Wiley. 1987. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature (Lond.)*. 329:506.
22. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1980. Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* 143:161.
23. Moore, K. W., B. Sher, Y. H. Sun, K. A. Eakle, and L. Hood. 1982. DNA sequence

of a gene encoding a BALB/c mouse L^d transplantation antigen. *Science (Wash. DC).* 215:679.

24. Kabat, E., T. Wu, M. Reid-Miller, H. Perry, and K. Gottesman. 1977. Sequences of Proteins of Immunological Interest, 4th ed. U.S. Department of Health and Human Services, Public Health Services, National Institutes of Health, Bethesda, MD.
25. Needleman, S., and C. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48:443.
26. Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA.* 78:454.
27. Eisenberg, D., E. Schwarz, M. Komaromy, and R. Wall. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* 179:125.
28. Eisenberg, D., R. Weiss, T. Terwilliger, and W. Wilcox. 1982. Hydrophobic moments and protein structure. *Faraday Symp. Chem. Soc.* 17:109.
29. Steinmetz, M., J. Frelinger, D. Fisher, T. Hunkapiller, D. Pereira, S. Weissman, H. Uehara, S. Nathenson, and L. Hood. 1981. Three cDNA clones encoding mouse transplantation antigens: homology to immunoglobulin genes. *Cell.* 24:125.
30. Chen, Y.-T., Y. Obata, E. Stockert, and L. Old. 1985. Thymus-Leukemia (TL) antigens of the mouse. Analysis of TL mRNA and TL cDNA from TL^+ and TL^- strains. *J. Exp. Med.* 162:1134.
31. Krayev, A., D. Kramerov, K. Skryabin, A. Ryskov, A. Bayev, and G. Georgiev. 1980. The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucleic Acids Res.* 8:1201.
32. Crews, S., J. Griffin, J. Huang, K. Calame, and L. Hood. 1981. A single V_H gene segment encodes the immune response to phosphoryl choline: Somatic mutation is correlated with the class of antibody. *Cell.* 25:59.
33. Kastern, W. 1985. Characterization of two class I major histocompatibility rat cDNA clones, one of which contains a premature termination codon. *Gene (Amst.).* 34:227.
34. Davis, N., and P. Model. 1985. An artificial anchor domain: hydrophobicity suffices to stop transfer. *Cell.* 41:607.
35. Klein, J., and F. Figueroa. 1986. The evolution of class I MHC genes. *Immunol. Today.* 7:41.
36. Hayashida, H., and T. Miyata. 1983. Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. *Proc. Natl. Acad. Sci. USA.* 80:2671.
37. Weiss, E. H., L. Golden, K. Fahrner, A. L. Mellor, J. J. Devlin, H. Bullman, H. Tiddens, H. Bud, and R. A. Flavell. 1984. Organization and evolution of the class I gene family in the major histocompatibility complex of the C57BL/10 mouse. *Nature (Lond.).* 310:650.
38. Devlin, J., E. Weiss, M. Paulson, and R. Flavell. 1985. Duplicated gene pairs and alleles of class I genes in the *Qa* region of the murine major histocompatibility complex: a comparison. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:3203.
39. Breathnach, R., and P. Chambon. 1981. Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50:349.
40. Mount, S. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10:459.
41. Tanford, C. 1980. The hydrophobic effect: Formation of Micelles and Biological Membranes. Wiley, New York.
42. Warren, G. 1981. Membrane proteins: structure and assembly. *In* Membrane Structure. J. B. Finean, and R. H. Michell, editors. Elsevier Science Publishers B. V., Amsterdam. 215–257.
43. Chou, P., and G. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. Relat. Areas Mol. Biol.* 47:45.
44. Karplus, P., and G. Schulz. 1985. Prediction of chain flexibility in proteins. *Naturwissen-*

schaften. 72:212.
45. Seegan, G., C. Smith, and V. Schumaker. 1979. Changes in quaternary structure of IgG upon reduction of the interheavy-chain disulfide bond. *Proc. Natl. Acad. Sci. USA.* 76:907.
46. Marquart, M., J. Deisenhofer, R. Huber, and W. Palm. 1980. Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0Å and 1.9Å resolution. *J. Mol. Biol.* 141:369.
47. Klein, M., N. Haeffner-Cavaillon, D. Isenman, C. Rivat, M. Navia, P. Davies, and K. Dorrington. 1981. Expression of biological effector functions by immunoglobulin G molecules lacking the hinge region. *Proc. Natl. Acad. Sci. USA.* 78:524.
48. Henderson, R., and P. N. T. Unwin. 1975. Three dimensional model of purple membrane obtained by electron microscopy. *Nature (Lond.).* 257:28.
49. Guidotti, G. 1977. The structure of intrinsic membrane proteins. *J. Supramol. Struct.* 7:489.
50. Davis, N., J. Boeke, and P. Model. 1985. Fine structure of a membrane anchor domain. *J. Mol. Biol.* 181:111.
51. Klapper, M. 1977. The independent distribution of amino acid near neighbor pairs into polypeptides. *Biochem. Biophys. Res. Commun.* 78:1018.
52. Clothia, C. 1976. The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* 105:1.
53. Malissen, M., T. Hunkapiller, and L. Hood. 1983. Nucleotide sequence of a light chain gene of the mouse *I-A* subregion: A^d_β. *Science (Wash. DC).* 221:750.
54. Tomita, M., and V. Marchesi. 1975. Amino acid sequence and oligosaccharide attachment sites of human erythrocyte glycophorin. *Proc. Natl. Acad. Sci. USA.* 72:2964.
55. Cosman, D., M. Kress, G. Khoury, and G. Jay. 1982. Tissue specific expression of an unusual *H-2* (class I)-related gene. *Proc. Natl. Acad. Sci. USA.* 79:4947.
56. Zuniga, M., and L. Hood. 1986. Clonal variation in cell surface display of an H-2 protein lacking a cytoplasmic tail. *J. Cell Biol.* 102:1.
57. Schepart, B., J. Woodward, M. Palmer, M. Macchi, P. Basta, E. McLaughlin-Taylor, and J. Frelinger. 1985. Expression in L cells of transfected class I genes from the mouse major histocompatibility complex. *Proc. Natl. Acad. Sci. USA.* 82:5505.
58. Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. Watson. 1983. Molecular Biology of the Cell. Garland Publishing, New York.
59. Sher, B., R. Nairn, J. E. Coligan, and L. Hood. 1985. DNA sequence of the mouse $H\text{-}2D^d$ transplantation antigen gene. *Proc. Natl. Acad. Sci. USA.* 82:1175.
60. Transy, C., S. R. Nash, B. David-Watine, M. Cochet, S. W. Hunt, L. E. Hood, and P. Kourilsky. 1987. A low polymorphic mouse *H-2* class I gene from the *Tla* complex is expressed in a broad variety of cell types. *J. Exp. Med.* 166:341.
61. Fisher, D. A., M. Pecht, and L. Hood. 1989. DNA sequence of a class I pseudogene from the *Tla* region of the murine MHC: recombination at a B2 Alu repetitive sequence. *J. Mol. Evol.* 28:306.
62. Headly, M., S. Hunt, K. Brorson, J. Andris, L. Hood, J. Forman, and P. Tucker. 1989. DNA sequence analysis of $D2^d$: a new *D*-region class I gene. *Immunogenetics.* 29:359.